PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/00, 27/12	A1	(11) International Publication Number: WO 97/14958 (43) International Publication Date: 24 April 1997 (24.04.97)

(21) International Application Number: PCT/GB96/02490

(22) International Filing Date: 11 October 1996 (11.10.96)

(30) Priority Data:
9521089.4 14 October 1995 (14.10.95) GB

(71) Applicant (*for all designated States except US*): AROMASCAN PLC [GB/GB]; Electra House, Electra Way, Crewe, Cheshire CW1 1WZ (GB).

(72) Inventor; and
(75) Inventor/Applicant (*for US only*): PERSAUD, Krishna, Chandra [GY/GB]; 65 Mersey Bank Avenue, Chorlton, Manchester M21 7NT (GB).

(74) Agents: McNEIGHT, David, Leslie et al.; McNeight & Lawrence, Regent House, Heaton Lane, Stockport, Cheshire SK4 1BS (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CLUSTER ANALYSIS

(57) Abstract

There is disclosed a concentration sensitive method for analysis of a plurality of outputs from chemical sensing device comprising the steps of: normalising said plurality of outputs; calculating at least one intensity output, said intensity output being related to the absolute magnitude of at least one of said plurality of outputs; and performing a cluster analysis of the plurality of normalised outputs and the intensity output, or outputs.

CLUSTER ANALYSIS

This invention relates to the use of cluster analysis in chemical sensing, in particular to the use of intensity data in such analyses in order to provide information regarding chemical concentrations.

In recent years there has been a great deal of interest in the field of gas sensing. [For the purposes of the present description, it is understood that 'gas sensing' comprises the detection of any chemical in the gas phase, including odours and volatile species]. One approach is to employ, within a single gas sensing device, an array of gas sensors which use semiconducting organic polymers (SOPs) as the active sensing material (see, for example, Persaud K C, Bartlett J G and Pelosi P, in 'Robots and Biological Systems : Towards a new bionics?', Eds. Darios P, Sandini G and Aebisher P, NATO ASI Series F : Computer and Systems Sciences 102 (1993) 579). Transduction is accomplished by measuring changes in the dc resistance of the sensors, these changes being induced by the absorption of gaseous species onto the SOPs.

The sensors are selected so as to exhibit differing but overlapping responses to a variety of gases, and therefore the output of an array of sensors is a pattern of response characteristic of the gas or gases detected. Since the number of sensors in an array is typically rather large - AromaScan plc manufacture devices having 20 and 32 sensor arrays - it can be said that these patterns are projected into multi-dimensional space of high order. Human vision is very good at recognising structural relationships within two and three dimensional space; however, in multi-dimensional space the perception of such relationships is extremely difficult. Therefore, in order for a human to examine complex multi-dimensional data, it is extremely useful to map such data from the high dimensional pattern space in which they are originally presented onto a low (two or three) dimensional pattern space.

There are numerous methods for performing the 'mapping' operation, which may comprise linear or non-linear algorithms. Linear mapping algorithms are used frequently for reasons of simplicity and generality. Such algorithms have been used in gas and odour classification as well as in chemical data classification in order to reduce multi-dimensional pattern space to two or three dimensional space. For gas recognition, Gardner *et al* (Gardner J W and Bartlett P N, Sensors and Actuators B 18-19 (1994) 221 and references therein) used a principal component analysis (PCA) method - a derivative of the Karhunen-Loeve (K-L) projection and one of the more powerful linear mapping techniques - to classify volatile chemicals by representing similar sets of data in characteristic 'clusters'. Ballantine Jr *et al* (Ballantine Jr. D S, Rose S L, Grate J W and Wohltjen H, Anal.Chem., 58 (1986) 3058) classified vapours using the PCA method and the (K-L) projection. The K-L projection was used in odour classification by Abe *et al* (Abe H, Kanaya S, Takahashi Y and Sasahi S-I, Analytica Chemica Acta 215 G988) 155) and Nakamoto *et al* (Nakamoto T, Fukuda A, Morizumi T and Asakura Y, Sensors and Actuators B, 3 (1991) 221) who investigated the odour of whisky data sets. Kowalski and Bender (Kowalski B R and Bender C F, J Amer.Chem. Soci., 95.3 (1973) 686) employed a similar linear mapping technique, with eigenvector projection, for displaying chemical data.

Non-linear mapping algorithms may be used when linear mapping is unable to preserve complex data structures - which is, in fact, commonly the case with 'real life' data. Non-linear techniques have complicated mathematical formulations compared to linear mapping, and are rarely used for gas classification. However, the responses of the array of sensors employed in the aforementioned AromaScan systems represent non-linear, multi-dimensional pattern structures, which (when normalised) contain the concentration independent pattern data sets describing different gases. In this instance non-linear mapping techniques are more applicable than linear techniques. It should be noted that truly concentration independent patterns are generated only when the concentration-response relationship is linear.

A particularly useful form of non-linear mapping is the algorithm of Sammon Jr. (Sammon Jr, JW, IEEE Trans. on Computers C-18 (1969) 401) and variations thereof, which represent highly effective methods of multivariate data analysis and clearly visualise multi-dimensional patterns onto two and three dimensional patterns. Various modifications to Sammon's algorithms have been proposed (see, for example, Kowalski and Bender, ibid; Nicemann H and Weiss J, IEEE Trans. on computers C-28 (1979) 142; Chang C L and Lee R C T, IEEE Trans. on System, man and cybernetics, (1973) 197; Pykett C E, Electron Lett., 14 (1978) 799; Biswas G, Jain A K and Dubes R C, IEEE Trans. on pattern analysis and machine intelligence, PAMI-3 (1981) 701) which are mainly concerned with reducing memory size and convergence time whilst remaining within the Sammon framework. Such considerations are no longer major problems due to the enormous recent advances in computer technology. Persaud *et al* (Hatfield J V, Neaves P, Hicks P J, Persaud K and Travers P, Sensors and Actuators B, 18-19 (1994) 221) have used the Sammon technique for vapour sensing applications in order to observe correlations between alcoholic data sets.

Since the mapping techniques described above result in 'clustering' of similar pattern types around characteristic two or three dimensional coordinates, the application of such techniques and the like will hereinafter be described as cluster analysis.

In the context of chemical sensing, prior art cluster analyses are essentially devoid of information regarding chemical concentration. This is because the cluster analysis is performed on patterns; raw sensor data - the intensity of which is related to chemical concentration - is scaled in an appropriate manner before cluster analysis. In instances where the concentration-sensor response relationship is non-linear, a pattern cluster will be skewed. In this sense the cluster analysis contains concentration information, but no direct use is made of absolute intensity data, and the effect is rather difficult to observe except at high concentrations/non-linearities.

The present invention overcomes the aforementioned difficulties by employing intensity information in cluster analyses in order to extract information on chemical concentration. Such fundamental information is frequently desirable, for instance, in the recognition of dangerously high levels of a toxic substance. It should be noted that whilst the invention is primarily directed towards the sensing of gaseous species, the approach is applicable to any area of chemical sensing where the sensing device produces a plurality of outputs which require some form of cluster analysis.

According to the invention there is provided a concentration sensitive method for analysis of a plurality of outputs from a chemical sensing device comprising the steps of :

normalising said plurality of outputs;

calculating at least one intensity output, each intensity output being related to the absolute magnitude of at least one of said plurality of outputs; and

performing a cluster analysis of the plurality of normalised outputs and the intensity output or outputs.

The intensity output, or outputs, may be weighted by a scaling factor.

The cluster analysis may comprise a non-linear mapping technique, and this technique may be the Sammon algorithm or a variant thereof.

A mathematical model of the results of the cluster analysis may be employed in order to derive quantitative concentration data.

There may be a single intensity output which is the mean of the moduli of the plurality of outputs.

There may be a plurality of intensity outputs wherein each of said intensity outputs comprises the absolute magnitude of an individual output.

The chemical sensing device may be a gas sensing device comprising at least one semiconducting organic polymer (SOP) based sensor, and the gas sensing device may further comprise an array of SOP based sensors wherein the outputs of the device correspond to changes in the dc resistance of said sensors.

Embodiments of concentration sensitive methods of analysis according to the invention will now be described with reference to the accompanying drawings, in which :

Figure 1 is a two dimensional cluster map;

and

Figure 2 is a graph of sensor response across an array of ten sensors.

The present invention is a concentration sensitive method of analysis of a plurality of outputs from a chemical sensing device comprising the steps of :

normalising said plurality of outputs;

calculating at least one intensity output, each intensity output being related to the absolute magnitudes of at least one of said plurality of outputs; and

performing a cluster analysis of the plurality of normalised outputs and the intensity output, or outputs.

Cluster analyses make no prior assumptions of the classes in which patterns belong, and apparent clustering of points is a matter for human judgement. In the field of chemical sensing, patterns generated by repeated exposure of a sensing device to a single compound of differing concentrations are identical if the concentration-output response relationship is linear. When a conventional cluster analysis is employed the

points coalesce into a single point or a closely grouped set of points, with the distances between the points representing experimental error. A cluster 10 of the latter type is shown in Figure 1.

However, it is often useful for the cluster analysis to reveal information on chemical concentration, e.g. two samples may be identical in composition but at different concentrations. A non-limiting example is provided by gas sensing devices of the type manufactured by AromaScan plc, which comprise an array of SOP sensors. Transduction is accomplished by measuring the changes in sensor dc resistances produced by exposure of the sensors to a gas or a mixture of gases. Figure 2 depicts a generalised response of an array of ten such sensors to a gas, the response comprising a plurality of outputs 20-38. The outputs 20-38 are recorded as $\Delta R/R$, the fractional change in resistance, where R is the base resistance of a sensor is clean air and ΔR is the change in resistance. It should be noted that an output may be negative. The absolute magnitude of a $\Delta R/R$ response (i.e. the modulus $|\Delta R/R|$) increases with increasing concentrations of the detected gas; one embodiment of the present invention utilises this fact by introducing to the cluster analysis an 'intensity' output which is related to the absolute magnitudes of the plurality of outputs 20-38. It is convenient to calculate the absolute mean intensity of the response.

Concentration independent patterns are produced by normalising the outputs 20-38 of the sensor array. The normalisation is performed by calculating the percentage fractional change in resistance for each sensor over the entire array. This given by equation (1) :

$$\frac{\frac{\Delta R}{R}}{\sum_{i=1}^{n} \left|\frac{\Delta R_i}{R_i}\right|} \times 100\% \qquad (1)$$

where n = 10 in the present example. The normalised outputs together with the intensity output are subjected to cluster analysis, the intensity output being scaled so that it is either comparable to the normal range of number present in pattern information or greater, so that the cluster analysis is biased towards intensity rather than pattern information. The scaling or weighting factor may be user determined.

As described earlier, the non-linear Sammon mapping technique, or variations thereof, represent a preferred class of cluster analysis in the case of SOP based sensor arrays for gas detection. However, other forms of cluster analysis (linear or non-linear) such as principal component analysis or variants such as factor analysis may also be applied. Indeed, such forms may prove preferable in other chemical sensing applications.

The results of a two dimensional analysis according to the present invention are displayed generally in Figure 1, which reveals that measurements of an odour at different concentrations thereof appear as a streak 12, the distance between two points being dependent on the difference in sample concentrations during the corresponding measurements.

In the above described embodiment a single intensity output, representing the absolute mean intensity of output response, is employed in the cluster analysis. An alternative approach, which is also within the scope of the invention , is to utilise a number of intensity outputs, each intensity output representing the absolute magnitude of a single selected sensor output. Thus a selected subset of the overall response to the sensor array may be employed in the cluster analysis. The intensity outputs may be scaled by suitable weighting factors.

It should be noted that generally when SOP based sensors of the type described above are exposed to a single gas, the concentration-response relationship is linear over a wide range of gas concentrations. However, when the array of sensors is exposed to a mixture of chemicals, the concentration response relationship may be non-linear, even if the mixture composition remains constant as the concentration varies. This phenomenon is due to competition for adsorption between compounds of differing binding affinities, since this competition is dependent on the concentrations of the compounds. At low concentrations compounds with the highest binding affinities are adsorbed onto the SOPs; and therefore the sensors are only responsive to these compounds. (The modulation of sensor resistance is due to - as yet not fully characterised - changes in SOP electronic structure and charge distribution caused by the adsorption of gases). As concentrations increase, compounds of lower binding affinity, begin to compete for binding. Therefore, normalised response patterns recorded at different concentrations will differ in appearance. As a result, cluster analysis gives rise to a streak, rather than a tight cluster. In this sense, the cluster analysis contains some information on chemical concentration, but any effect is difficult to observe at low concentrations. The use of intensity data in the cluster analysis results in concentration dependent mapping in which it is easy to visually distinguish one point from another on the basis of concentration.

A further aspect of the present invention is the extraction of quantitative concentration data from the results of the cluster analysis. Since the distances between points are proportional to concentration, it is possible to apply an appropriate mathematical model (such as a polynomial fit), to the data in order to interpolate or extrapolate unknown patterns and thereby extract concentrations.

It will be appreciated that it is not intended to limit the invention to the above examples only, many variations, such as might readily occur to one skilled in the art, being possible without departing from the scope thereof. For instance, the plurality

of outputs used in the cluster analysis need not emanate from an array of sensors. UK Patent GB 2 203 553 B discloses a SOP based sensor used in conjunction with an ac transduction technique. In this instance, it may be desirable to measure changes in impedance characteristics at a plurality of ac frequencies : in this way, a single sensor may provide the plurality of outputs. The outputs of arrays of chemical sensors used to monitor liquid analytes may also be amenable to the cluster analysis described herein.

CLAIMS

1. A concentration sensitive method for analysis of a plurality of outputs from chemical sensing device comprising the steps of :

normalising said plurality of outputs;

calculating at least one intensity output, said intensity output being related to the absolute magnitude of at least one of said plurality of outputs; and

performing a cluster analysis of the plurality of normalised outputs and the intensity output, or outputs.

2. A concentration sensitive method according to claim 1 in which the intensity output, or outputs, is weighted by a scaling factor.

3. A concentration sensitive method according to claim 1 or claim 2 in which the cluster analysis comprises a non-linear mapping technique.

4. A concentration sensitive method according to claim 3 in which the non-linear mapping technique is the Sammon algorithm or a variant thereof.

5. A concentration sensitive method according to any of the previous claims in which a mathematical model of the results of the cluster analysis is employed to derive quantitative concentration data.

6. A concentration sensitive method according to any of the previous claims in which a single intensity output is calculated, said intensity output being the mean of the moduli of the plurality of outputs.

7. A concentration sensitive method according to any of claims 1 - 5 in which a plurality of intensity outputs are calculated, each of said intensity outputs comprising the absolute magnitude of an individual output.

8. A concentration sensitive method according to any of the previous claims in which the chemical sensing device is a gas sensing device comprising at least one semiconducting organic polymer based sensor.

9. A concentration sensitive method according to claim 8 in which the gas sensing device comprises an array of sensors and the outputs of the device correspond to changes in the dc resistance of said sensors.

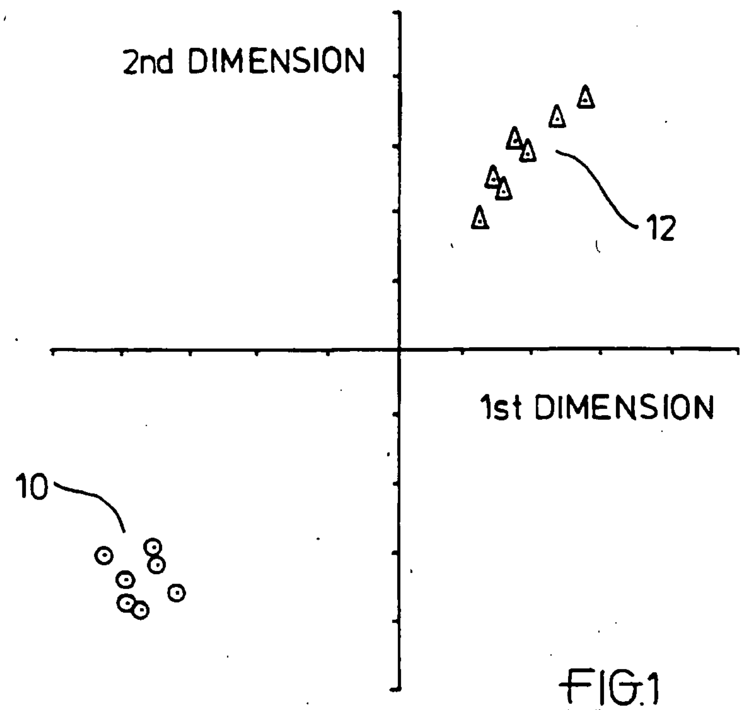
2nd DIMENSION
12
1st DIMENSION
10

FIG.1

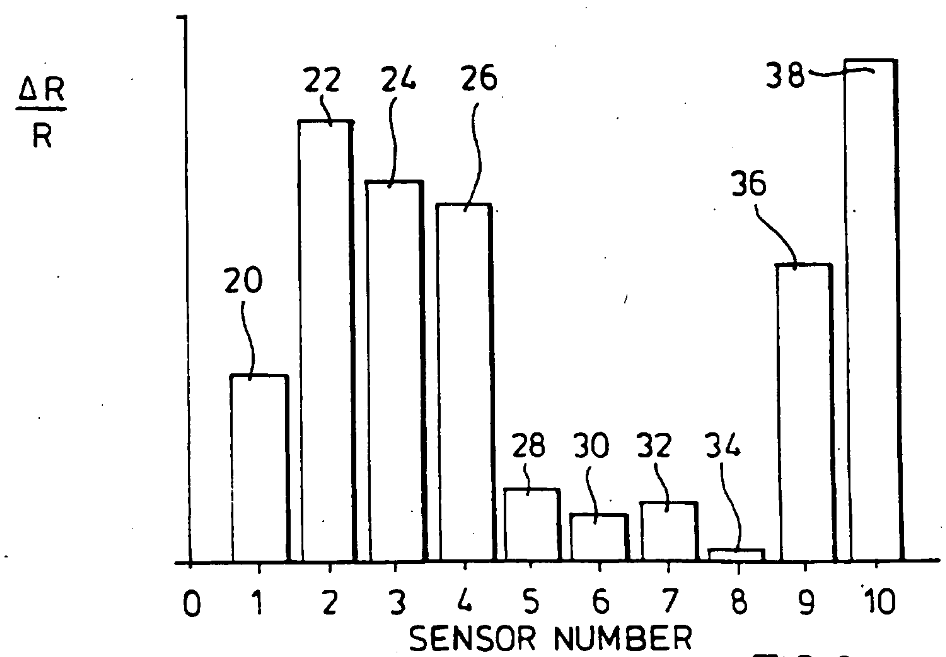
ΔR
R
22
24
26
38
36
20
28
30
32
34
0
1
2
3
4
5
6
7
8
9
10
SENSOR NUMBER

FIG.2

INTERNATIONAL SEARCH REPORT

Inter. nal Application No
PCT/GB 96/02490

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G01N33/00 G01N27/12

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	SENSORS AND ACTUATORS B, vol. 4, 1991, LAUSANNE CH, pages 109-115, XP002024604 J. W. GARDNER: see the whole document ---	1 2-5
Y	JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 95.3, 1973, pages 686-693, XP000615433 B. R. KOWALSKI ET AL.: cited in the application see page 691, left-hand column ---	2-5

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
5 February 1997	- 4. 03. 97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Authorized officer Brison, O

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 96/02490

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TECHNISCHES MESSEN TM 1982 - 1988 INCOMPLETE, vol. 62, no. 4, 1 April 1995, MÜNCHEN, DE, pages 166-172, XP000516380 HORNER G: "SIGNALVERARBEITUNG BEI CHEMOSENSORARRAYS" see page 168 - page 169	1
A	ANALYTICAL CHEMISTRY, vol. 58, no. 4, April 1986, WASHINGTON, DC, US, pages 860-866, XP002024605 J. R. STETTER ET AL.: see page 861, right-hand column - page 862, left-hand column	1-4
X	WO 86 01599 A (COGENT LTD) 13 March 1986 see abstract	8,9
X	SENSORS AND ACTUATORS B, vol. 18-19, 1994, LAUSANNE CH, pages 211-220, XP000615104 J. W. GARDNER ET AL.: see page 214 - page 218	1
A	US 4 638 443 A (KANEYASU MASAYOSHI ET AL) 20 January 1987 see column 5, line 22 - column 8, line 6	1
A	ANALYTICA CHIMICA ACTA, vol. 215, 1988, AMSTERDAM, NL, pages 155-168, XP002024606 H. ABE ET AL.: cited in the application see page 161	1
A,P	WO 95 32420 A (QUEST INT ;BEHAN JOHN MARTIN (GB); PERRING KEITH DOUGLAS (GB); WIL) 30 November 1995 see page 4 - page 11	1,8,9

1

INTERNATIONAL SEARCH REPORT

...formation on patent family members

Intern al Application No
PCT/GB 96/02490

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-8601599	13-03-86	EP-A- 0226584	01-07-87
US-A-4638443	20-01-87	JP-B- 4039029	26-06-92
		JP-A- 59153159	01-09-84
WO-A-9532420	30-11-95	NONE	